

LEAD SCORE CASE STUDY

Submitted by:
Asma Thasneem

Problem Statement

X Education sells online courses to industry professionals.

- ▶ X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- ▶ To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- ▶ If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Business Objective:

- ▶ X education wants to know most promising leads.
- ▶ For that they want to build a Model which identifies the hot leads.
- ▶ Deployment of the model for the future use

Solution Methodology

Data cleaning and data manipulation.

- ▶ Check and handle NA values and missing values.
- ▶ Drop columns, if it contains large amount of missing values and not useful for the analysis.
- ▶ Imputation of the values, if necessary.
- ▶ Check and handle outliers in data.

▶ EDA

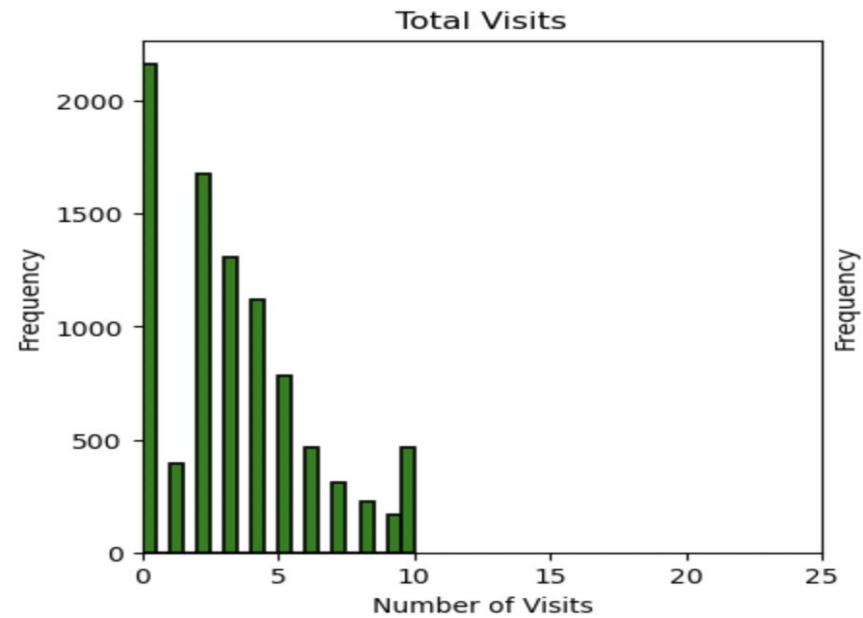
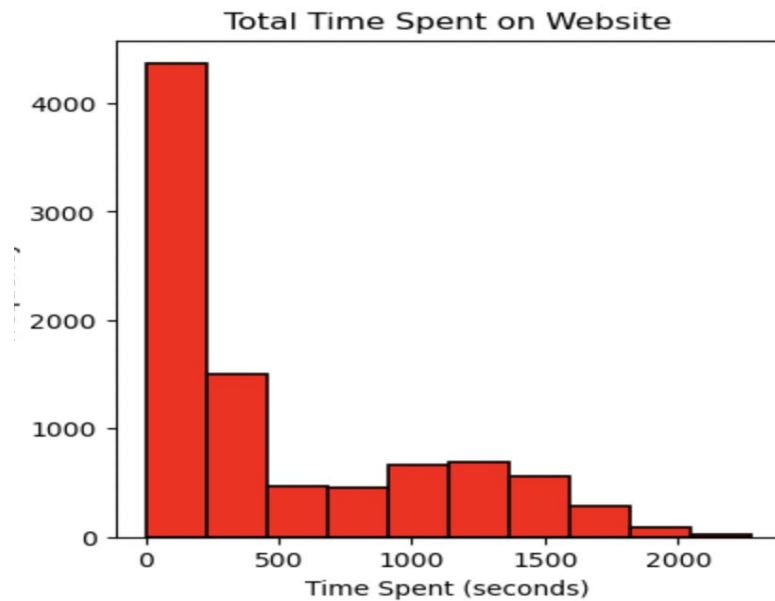
1. Univariate data analysis: value count, distribution of variable etc.
2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.

- ▶ Feature Scaling & Dummy Variables and encoding of the data.
- ▶ Classification technique: logistic regression used for the model building and prediction.
- ▶ Validation of the model.
- ▶ Model presentation.
- ▶ Conclusions and recommendations

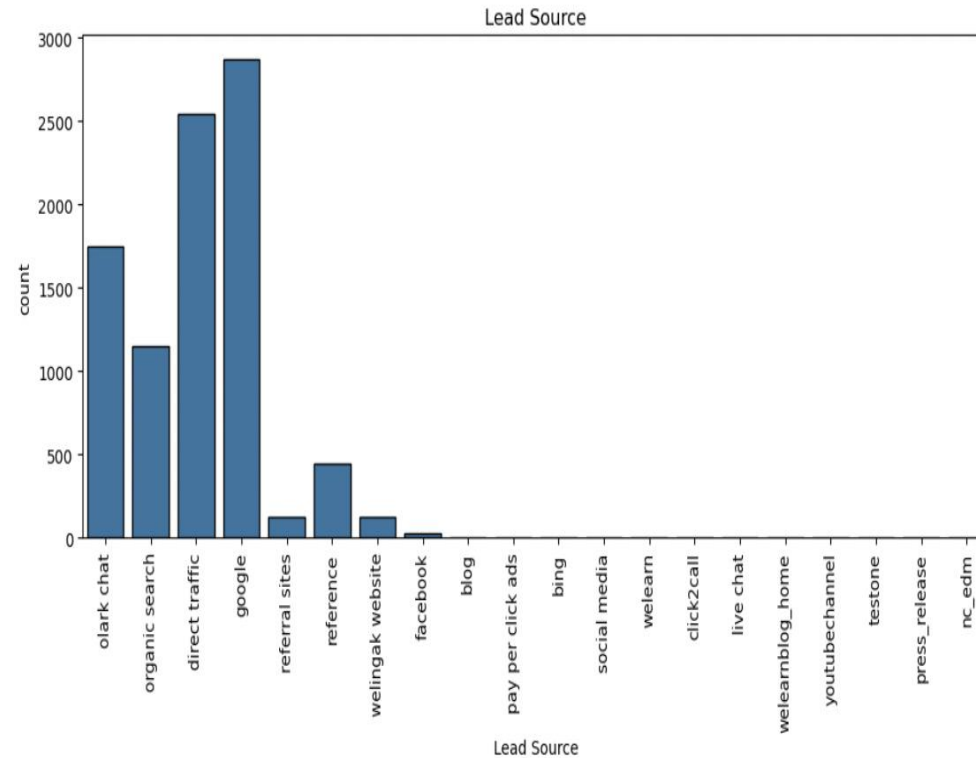
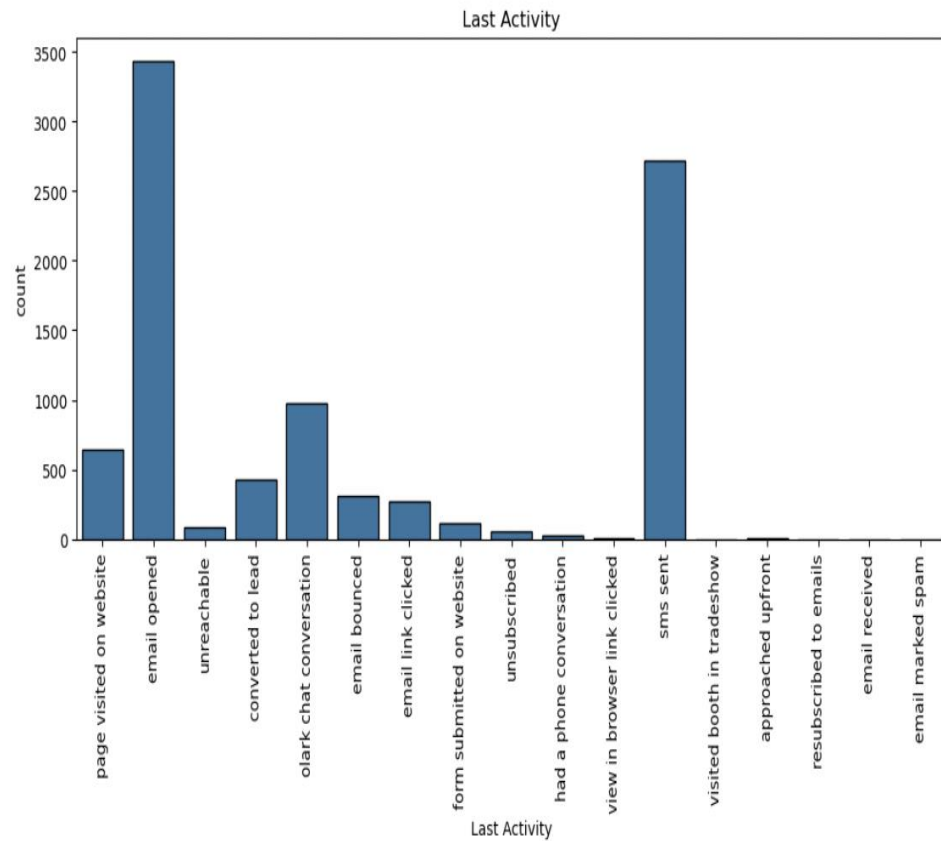
Data Manipulation

- ▶ Removing the “Prospect ID” and “Lead Number” which is not necessary for the analysis
- ▶ There are a few columns in which there is a level called select which is converted as null value.
- ▶ Dropping the columns having more than 30% null values.
- ▶ After checking for the value counts for some of the object type variables, we find some of the features has no enough variance, which we have dropped, the features are: “Do Not Call”, “What matters most to you in choosing course”, “Search”, “Newspaper Article”, “X Education Forums”, “Newspaper”, “Digital Advertisement” etc.

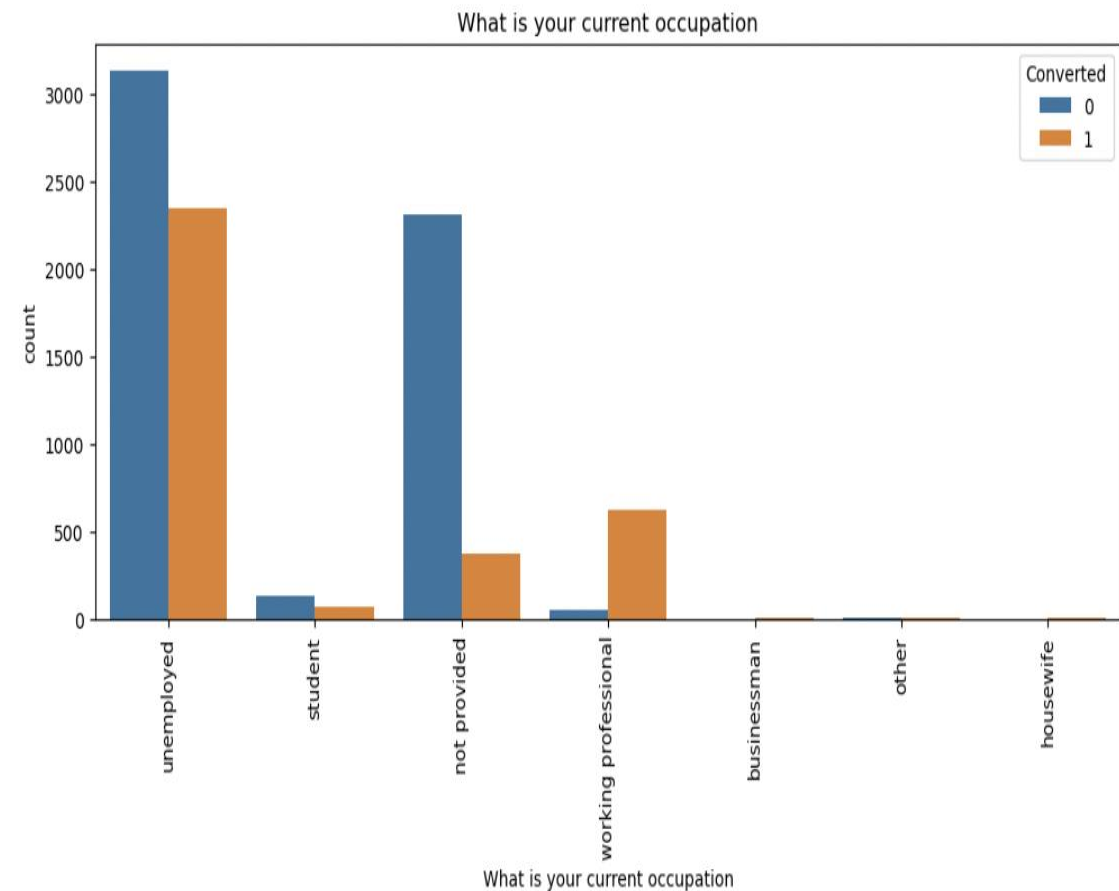
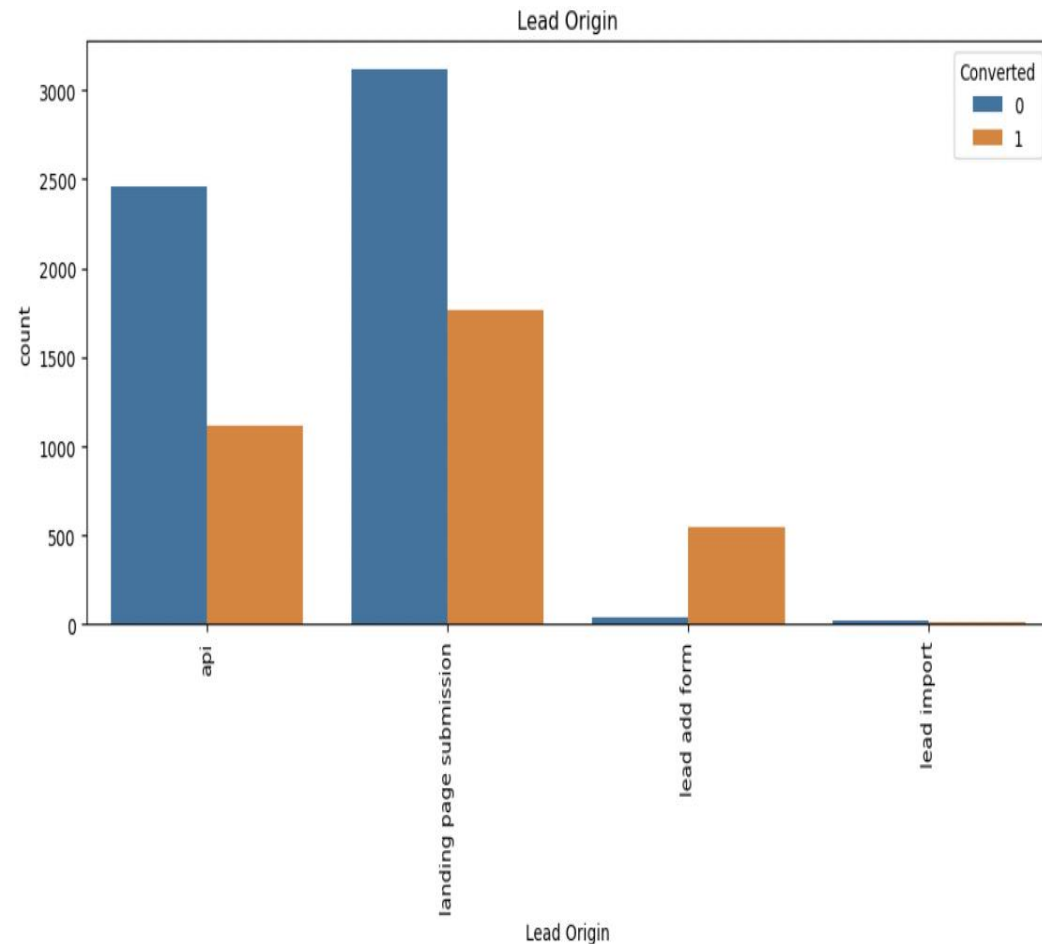
- EDA - Univariate Analysis for Numerical Variables



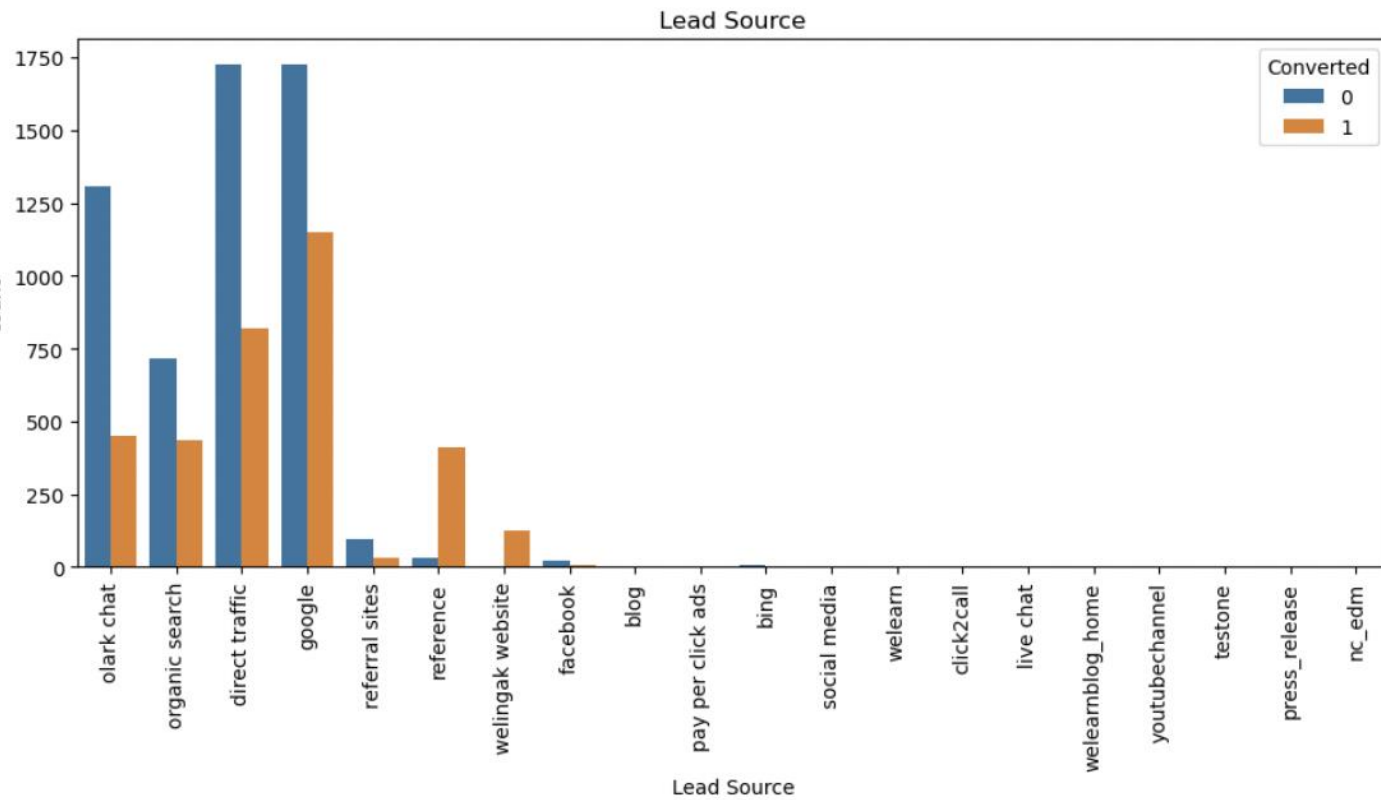
Univariate Analysis for Categorical Variables



Categorical Variable Relation



- In Lead origin most number of converted leads are on lead add form.
- In Current Occupation most number of converted leads are on working professional.



In lead source the leads through reference have high probability to convert.

Data Conversion

- ▶ Numerical Variables are Normalized
- ▶ Dummy Variables are created for object type variables

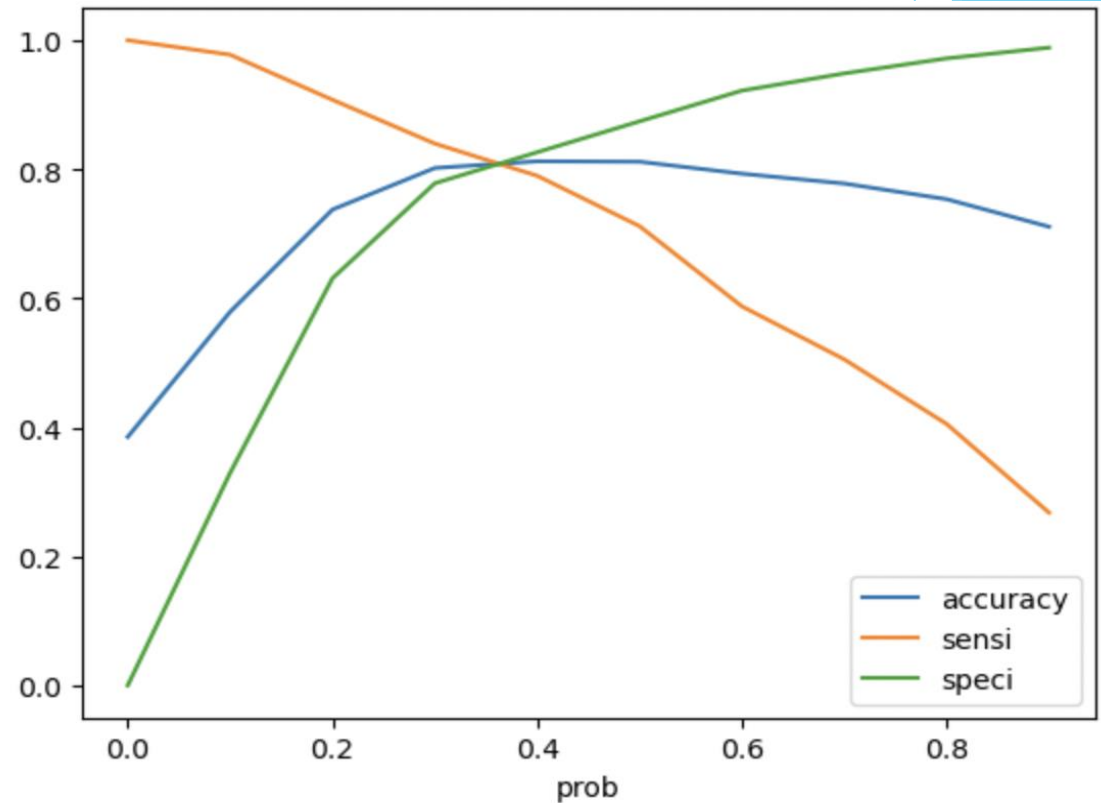
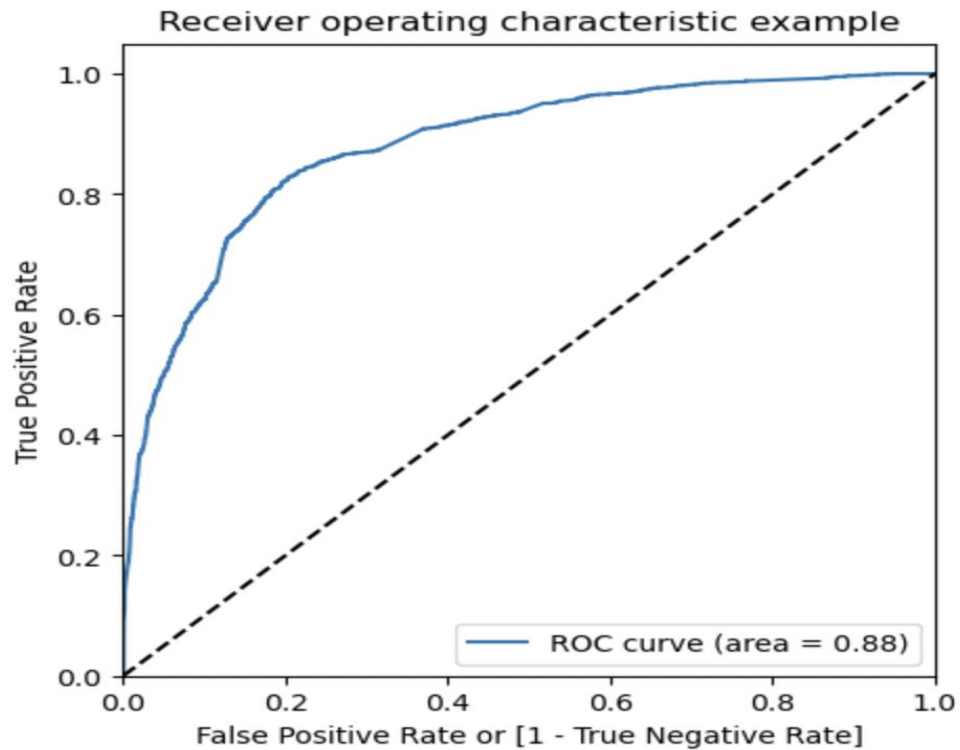
Model Building

- ▶ Splitting the Data into Train and Test Sets
- ▶ The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- ▶ Use RFE for Feature Selection
- ▶ Running RFE with 15 variables as output
- ▶ Building Model by removing the variable whose p- value is greater than 0.05 and vif value is greater than 5

Model Evaluation

- ▶ Evaluating confusion matrix
- ▶ Rather than accuracy evaluating other matrices such as sensitivity , Specificity , Precision and recall for various cut offs.

ROC Curve



Finding Optimal Cut off Point

- Optimal cut off probability is that probability where we get balanced sensitivity and specificity.
- From the second graph it is visible that the optimal cut off is at 0.39

Conclusion

The top factors that Positively Impact Lead Conversion are

- 1.Total Time Spent on Website
- 2.When their current occupation is a working professional.
- 3.When the lead origin is add form
- 4.Last Activity - Had a Phone Conversation

The top factors that Negatively Impact Lead Conversion

- 1.Do Not Email - Yes
- 2.Last Notable Activity - Email Opened
- 3.Last Notable Activity - Page Visited on Website

Recommendation

Based on the significant variables identified in the lead scoring case study, here are some actionable recommendations to improve lead conversion:

- ▶ Prioritize Leads with High Website Engagement
- ▶ Target Working Professionals
- ▶ Focus on Leads from the Add Form
- ▶ Leverage Email Engagement Data
- ▶ Optimize Follow-Up Strategies